

The listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (previously presented) A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

i) encoding two or more biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;

ii) selecting at least two substrings from said initial character strings;

iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adding the product strings to a data structure to populate a data structure of product strings;

v) determining sequence identities of the product strings relative to at least one initial character string; and

vi) selecting one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character string.

2. (previously presented) The method of claim 1, wherein said encoding comprises encoding two or more nucleic acid sequences into said character strings.

3. (previously presented) The method of claim 2, wherein said two or more nucleic acid sequences comprise a nucleic acid sequence encoding a naturally occurring protein.

4. (previously presented) The method of claim 1, wherein said encoding comprises encoding two or more amino acid sequences into said character strings.

5. (previously presented) The method of claim 4, wherein said two or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

6. (previously presented) The method of claim 1, wherein said initial character strings have at least 30% sequence identity with each other.

7. (previously presented) The method of claim 1, wherein said selecting in (ii) comprises selecting at least one substring from an initial character string such that the ends of said substring occur in string regions of about 3 to about 20 characters in the initial character string that have higher sequence identity with the corresponding region of another of said initial character strings than the overall sequence identity between the two initial character strings.

8. (previously presented) The method of claim 1, wherein said selecting in (ii) comprises selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

9. (canceled)

10. (previously presented) The method of claim 1, wherein said selecting in (ii) comprises aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

11. (canceled)

12. (previously presented) The method of claim 1, wherein said method further comprises randomly altering one or more characters of said initial or product character strings.

13. (original) The method of claim 12, wherein said method further comprises randomly selecting and altering one or more occurrences of a particular preselected character in said character strings.

14. (previously presented) The method of claim 1, wherein said encoding, selecting, or concatenating is performed on an internet site.

15. (previously presented) The method of claim 1, wherein said encoding, selecting, or concatenating is performed on a server.

16. (previously presented) The method of claim 1, wherein said encoding, selecting, or concatenating is performed on a client linked to a network.

17. (previously presented) A computer program product on a computer readable media comprising computer code that:

- i) encodes two or more biological molecules into initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;
- ii) selects at least two initial substrings from said character strings;
- iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;
- iv) adds the product strings to a data structure to populate a data structure of product strings;
- v) determines sequence identities of the product strings relative to at least one initial character string; and
- vi) selects one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character string.

18. (previously presented) The computer program product of claim 17, wherein said two or more biological molecules are nucleic acid sequences encoding naturally occurring proteins.

19. (previously presented) The computer program product of claim 17, wherein said two or more biological molecules are nucleic acid sequences encoding naturally occurring proteins.

20. (previously presented) The computer program product of claim 17, wherein said two or more biological molecules are amino acid sequences.

21. (previously presented) The computer program product of claim 17, wherein said initial character strings have at least 30% sequence identity.

22. (previously presented) The computer program product of claim 17, wherein said computer code selects in (ii) at least one substring from an initial character string such that the ends of said substring occur in string regions of about three to about twenty characters in the initial character string that have higher sequence identity with a corresponding region of another of said initial character strings than the overall sequence identity between the two initial character substrings.

23. (previously presented) The computer program product of claim 17, wherein said computer code selects substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

24. (canceled)

25. (previously presented) The computer program product of claim 17, wherein the computer code selects substrings by aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the character strings, and selecting a character that is a member of an aligned pair for the end of one substring.

26. (canceled)

27. (previously presented) The computer program product of claim 17, wherein said computer code additionally randomly alters one or more characters of said character strings.

28. (previously presented) The computer program product of claim 27, wherein said computer code additionally randomly selects and alters one or more occurrences of a particular preselected character in said character strings.

29. (previously presented) The computer program product of claim 17, wherein said computer code is stored on media selected from the group consisting of magnetic media, optical media, and optomagnetic media.

30. (previously presented) The computer program product of claim 17, wherein said computer code is in dynamic or static memory of a computer.

31-44. (canceled)

45. (previously presented) The method of claim 1, wherein the initial character strings of (i) are related.

46. (canceled)

47. (previously presented) The method of claim 1, further comprising determining a computationally predicted property for molecules represented by the product strings.

48. (previously presented) The method of claim 1, wherein the molecules represented by the product strings are made in parallel in an array of vessels.

49. (previously presented) The method of claim 1, wherein the molecules represented by the product strings are made by assembly of oligonucleotides.

50. (canceled)

51. (previously presented) The computer program product of claim 17, wherein the initial character strings of (i) are related.

52. (previously presented) The computer program product of claim 17, wherein the code instructs physical screening of the molecule(s) represented by the product strings for one or more desired properties.

53. (previously presented) The computer program product of claim 17, wherein the code instructs determination of a computationally predicted property for molecules represented by the product strings.

54. (previously presented) The computer program product of claim 17, wherein the molecules represented by the product strings are made in parallel in an array of vessels.

55. (previously presented) The computer program product of claim 17, wherein the molecules represented by the product strings are made by assembly of oligonucleotides.

56. (currently amended) The computer program product of claim 17, wherein the code tests members of the data structure of product strings for a particular property and determines sequence differences responsible for differences in the particular property using multi-variate analysis.

57. (previously presented) A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

- i) encoding two or more related biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;
- ii) selecting at least two substrings from said initial character strings;
- iii) concatenating said substrings to form one or more product strings;

- iv) adding the product strings to a data structure to populate a data structure of product strings; and
- v) determining whether the product strings have at least a predefined measure of similarity with at least one initial character string; and
- vi) selecting one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings determined to have greater than the predefined value of sequence identity with at least one initial string.

58. (previously presented) The method of claim 1, wherein the one or more product strings of (vi) have greater than 50% sequence identity with the at least one initial character string..

59. (previously presented) The method of claim 1, wherein the one or more product strings of (vi) have greater than 75% sequence identity with the at least one initial character string.

60. (previously presented) The method of claim 1, wherein the one or more product strings of (vi) have greater than 85% sequence identity with the at least one initial character string.

61. (previously presented) The method of claim 1, wherein the one or more product strings of (vi) have greater than 90% sequence identity with the at least one initial character string.

62. (previously presented) The method of claim 1, wherein the one or more product strings of (vi) have greater than 95% sequence identity with the at least one initial character string.

63. (previously presented) The computer program product of claim 17, wherein the one or more product strings of (vi) having greater than 50% sequence identity with the at least one initial character string.

64. (previously presented) The computer program product of claim 17, wherein the one or more product strings of (vi) having greater than 75% sequence identity with the at least one initial character string.

65. (previously presented) The computer program product of claim 17, wherein the one or more product strings of (vi) having greater than 95% sequence identity with the at least one initial character string.

66. (previously presented) A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

- i) encoding two or more biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;
- ii) selecting at least two substrings from said initial character strings;
- iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;
- iv) adding the product strings to a data structure to populate a data structure of product strings;
- v) providing an alignment of the product strings; and
- vi) selecting one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character string.

67. (previously presented) The method of claim 66, wherein said encoding comprises encoding two or more amino acid sequences into said character strings, and wherein said two or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

68. (previously presented) The method of claim 66, wherein said initial character strings have at least 30% sequence identity with each other.

69. (previously presented) The method of claim 66, wherein said selecting in (ii) comprises selecting at least one substring from an initial character string such that the ends of said substring occur in string regions of about 3 to about 20 characters in the initial character string that have higher sequence identity with the corresponding region of another of said initial character strings than the overall sequence identity between the two initial character strings.

70. (previously presented) The method of claim 66, wherein said selecting in (ii) comprises selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

71. (previously presented) The method of claim 66, wherein said selecting in (ii) comprises aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

72. (previously presented) The method of claim 66, wherein said method further comprises randomly altering one or more characters of said initial or product character strings.

73. (previously presented) The method of claim 66, wherein the one or more product strings of (vi) having greater than 50% sequence identity with the at least one initial character string.

74. (previously presented) The method of claim 66, wherein the one or more product strings of (vi) having greater than 75% sequence identity with the at least one initial character string.

75. (previously presented) The method of claim 66, wherein the one or more product strings of (vi) having greater than 85% sequence identity with the at least one initial character string.

76. (previously presented) The method of claim 66, wherein the one or more product strings of (vi) having greater than 90% sequence identity with the at least one initial character string.

77. (previously presented) The method of claim 66, wherein the one or more product strings of (vi) having greater than 95% sequence identity with the at least one initial character string.

78. (previously presented) A computer program product on a computer readable media comprising computer code that:

- i) encodes two or more biological molecules into initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;
- ii) selects at least two initial substrings from said character strings;
- iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;

- iv) adds the product strings to a data structure to populate a data structure of product strings;
- v) provides an alignment of the product strings; and
- vi) selects one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character string.

79. (previously presented) The computer program product of claim 78, wherein said computer code encodes two or more amino acid sequences into said character strings, and wherein said two or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

80. (previously presented) The computer program product of claim 78, wherein said initial character strings have at least 30% sequence identity with each other.

81. (previously presented) The computer program product of claim 78, wherein said computer code selects in (ii) at least one substring from an initial character string such that the ends of said substring occur in string regions of about three to about twenty characters in the initial character string that have higher sequence identity with a corresponding region of another of said initial character strings than the overall sequence identity between the two initial character substrings.

82. (previously presented) The computer program product of claim 78, wherein said computer code selects in (ii) by selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

83. (previously presented) The computer program product of claim 78, wherein said computer code selects in (ii) by aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

84. (previously presented) The computer program product of claim 78, wherein said computer code further randomly alters one or more characters of said initial or product character strings.

85. (previously presented) The computer program product of claim 78, wherein the one or more product strings of (vi) having greater than 50% sequence identity with the at least one initial character string.

86. (previously presented) The computer program product of claim 78, wherein the one or more product strings of (vi) having greater than 75% sequence identity with the at least one initial character string.

87. (previously presented) The computer program product of claim 78, wherein the one or more product strings of (vi) having greater than 85% sequence identity with the at least one initial character string.

88. (previously presented) The computer program product of claim 78, wherein the one or more product strings of (vi) having greater than 90% sequence identity with the at least one initial character string.

89. (previously presented) The computer program product of claim 78, wherein the one or more product strings of (vi) having greater than 95% sequence identity with the at least one initial character string.

90. (previously presented) A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

- i) encoding two or more naturally occurring biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;
- ii) selecting at least two substrings from said initial character strings;
- iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;
- iv) adding the product strings to a data structure to populate a data structure of product strings; and
- v) selecting one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character string.

91. (previously presented) The method of claim 90, wherein said encoding comprises encoding two or more nucleic acid sequences into said character strings.

92. (previously presented) The method of claim 90, wherein said encoding comprises encoding two or more amino acid sequences into said character strings, and wherein said two or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

93. (previously presented) The method of claim 90, wherein said initial character strings have at least 30% sequence identity with each other.

94. (previously presented) The method of claim 90, wherein said selecting in (ii) comprises selecting at least one substring from an initial character string such that the ends of said substring occur in string regions of about 3 to about 20 characters in the initial character string that have higher sequence identity with the corresponding region of another of said initial character strings than the overall sequence identity between the two initial character strings.

95. (previously presented) The method of claim 90, wherein said selecting in (ii) comprises selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

96. (previously presented) The method of claim 90, wherein said selecting in (ii) comprises aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

97. (previously presented) The method of claim 90, wherein said method further comprises randomly altering one or more characters of said initial or product character strings.

98. (previously presented) The method of claim 90, wherein the one or more product strings of (v) having greater than 50% sequence identity with the at least one initial character string.

99. (previously presented) The method of claim 90, wherein the one or more product strings of (v) having greater than 75% sequence identity with the at least one initial character string.

100. (previously presented) The method of claim 90, wherein the one or more product strings of (v) having greater than 85% sequence identity with the at least one initial character string.

101. (previously presented) The method of claim 90, wherein the one or more product strings of (v) having greater than 90% sequence identity with the at least one initial character string.

102. (previously presented) The method of claim 90, wherein the one or more product strings of (v) having greater than 95% sequence identity with the at least one initial character string.

103. (previously presented) A computer program product on a computer readable media comprising computer code that:

- i) encodes two or more naturally occurring biological molecules into initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;
- ii) selects at least two initial substrings from said character strings;
- iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;
- iv) adds the product strings to a data structure to populate a data structure of product strings; and
- v) selects one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character_string.

104. (previously presented) The computer program product of claim 103, wherein said computer code encodes by encoding two or more nucleic acid sequences into said character strings.

105. (previously presented) The computer program product of claim 103, wherein said computer code encodes two or more amino acid sequences into said character strings, and wherein said two or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

106. (previously presented) The computer program product of claim 103, wherein said initial character strings have at least 30% sequence identity with each other.

107. (previously presented) The computer program product of claim 103, wherein said computer code selects in (ii) at least one substring from an initial character string such that the

ends of said substring occur in string regions of about three to about twenty characters in the initial character string that have higher sequence identity with a corresponding region of another of said initial character strings than the overall sequence identity between the two initial character substrings.

108. (previously presented) The computer program product of claim 103, wherein said computer code selects in (ii) by selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

109. (previously presented) The computer program product of claim 103, wherein said computer code selects in (ii) by aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

110. (previously presented) The computer program product of claim 103, wherein said computer code further randomly alters one or more characters of said initial or product character strings.

111. (previously presented) The computer program product of claim 103, wherein the one or more product strings of (v) having greater than 50% sequence identity with the at least one initial character string.

112. (previously presented) The computer program product of claim 103, wherein the one or more product strings of (v) having greater than 75% sequence identity with the at least one initial character string.

113. (previously presented) The computer program product of claim 103, wherein the one or more product strings of (v) having greater than 85% sequence identity with the at least one initial character string.

114. (previously presented) The computer program product of claim 103, wherein the one or more product strings of (v) having greater than 90% sequence identity with the at least one initial character string.

115. (previously presented) The computer program product of claim 103, wherein the one or more product strings of (v) having greater than 95% sequence identity with the at least one initial character string.

116. (previously presented) A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

- i) encoding two or more biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;
- ii) selecting at least two substrings from said initial character strings;
- iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;
- iv) adding the product strings to a data structure to populate a data structure of product strings;
- v) obtaining one or more computationally predicted properties for the product strings in the data structure; and
- vi) selecting one or more product biological molecules for production on the basis of the one or more computationally predicted properties.

117. (previously presented) The method of claim 116, wherein the computationally predicted properties comprise one or more of a maximum or minimum molecular weight, a maximum or minimum free energy, a maximum or minimum contact surface with a target molecule or surface, a specified net charge, a predicted pK, a predicted pI, a binding avidity, secondary form, and tertiary form.

118. (previously presented) The method of claim 116, wherein said encoding comprises encoding two or more amino acid sequences into said character strings.

119. (previously presented) The method of claim 116, wherein said selecting in (ii) comprises aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

120. (previously presented) The method of claim 116, wherein said method further comprises randomly altering one or more characters of said initial or product character strings.

121. (previously presented) The method of claim 116, wherein the one or more product biological molecules of (vi) having greater than 50% sequence identity with the at least one initial character string.

122. (previously presented) The method of claim 116, wherein the one or more product biological molecules of (vi) having greater than 75% sequence identity with the at least one initial character string.

123. (previously presented) The method of claim 116, wherein the one or more product biological molecules of (vi) having greater than 90% sequence identity with the at least one initial character string.

124. (previously presented) A computer program product on a computer readable media comprising computer code that:

- i) encodes two or more biological molecules into initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;
- ii) selects at least two initial substrings from said character strings;
- iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;
- iv) adds the product strings to a data structure to populate a data structure of product strings;
- v) obtains one or more computationally predicted properties for the product strings in the data structure; and
- vi) selects one or more product biological molecules for production on the basis of the one or more computationally predicted properties.

125. (previously presented) The computer program product of claim 124, wherein the computationally predicted properties comprise one or more of a maximum or minimum molecular weight, a maximum or minimum free energy, a maximum or minimum contact surface with a target molecule or surface, a specified net charge, a predicted pK, a predicted pI, a binding avidity, secondary form, and tertiary form.

126. (previously presented) The computer program product of claim 124, wherein the computer code encodes in (i) by encoding two or more amino acid sequences into said character strings.

127. (previously presented) The computer program product of claim 124, wherein the computer code selects in (ii) by aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

128. (previously presented) The computer program product of claim 124, wherein the computer code further randomly alters one or more characters of said initial or product character strings.

129. (previously presented) The computer program product of claim 124, wherein the one or more product biological molecules of (vi) having greater than 50% sequence identity with the at least one initial character string.

130. (previously presented) The computer program product of claim 124, wherein the one or more product biological molecules of (vi) having greater than 75% sequence identity with the at least one initial character string.

131. (previously presented) The computer program product of claim 124, wherein the one or more product biological molecules of (vi) having greater than 90% sequence identity with the at least one initial character string.

Please add the following new claims:

132. (new) The method of claim 1, wherein adding the product strings to a data structure comprises adding more than one product strings to the data structure.

133. (new) The method of claim 1, wherein selecting at least two substrings from said initial character strings comprises random substring selection.

134. (new) The method of claim 1, wherein selecting at least two substrings from said initial character strings comprises uniform substring selection.

135. (new) The method of claim 1, wherein selecting at least two substrings from said initial character strings comprises motif-based selection.

136. (new) The method of claim 1, wherein selecting at least two substrings from said initial character strings comprises alignment-based selection.

137. (new) The method of claim 1, wherein selecting at least two substrings from said initial character strings comprises frequency-biased selection.

138. (new) The computer program product of claim 17, wherein the computer code adds the product strings to a data structure by adding more than one product strings to the data structure.

139. (new) The computer program product of claim 17, wherein the computer code selects at least two substrings from said initial character strings by a random substring selection.

140. (new) The computer program product of claim 17, wherein the computer code selects at least two substrings from said initial character strings by a uniform substring selection.

141. (new) The computer program product of claim 17, wherein the computer code selects at least two substrings from said initial character strings by a motif-based selection.

142. (new) The computer program product of claim 17, wherein the computer code selects at least two substrings from said initial character strings by an alignment-based selection.

143. (new) The computer program product of claim 17, wherein the computer code selects at least two substrings from said initial character strings by a frequency-biased selection.

144. (new) The method of claim 66, wherein adding the product strings to a data structure comprises adding more than one product strings to the data structure.

145. (new) The method of claim 66, wherein selecting at least two substrings from said initial character strings comprises random substring selection.

146. (new) The method of claim 66, wherein selecting at least two substrings from said initial character strings comprises uniform substring selection.

147. (new) The method of claim 66, wherein selecting at least two substrings from said initial character strings comprises motif-based selection.

148. (new) The method of claim 66, wherein selecting at least two substrings from said initial character strings comprises alignment-based selection.

149. (new) The method of claim 66, wherein selecting at least two substrings from said initial character strings comprises frequency-biased selection.

150. (new) The computer program product of claim 78, wherein the computer code adds the product strings to a data structure by adding more than one product strings to the data structure.

151. (new) The computer program product of claim 78, wherein the computer code selects at least two substrings from said initial character strings by a random substring selection.

152. (new) The computer program product of claim 78, wherein the computer code selects at least two substrings from said initial character strings by a uniform substring selection.

153. (new) The computer program product of claim 78, wherein the computer code selects at least two substrings from said initial character strings by a motif-based selection.

154. (new) The computer program product of claim 78, wherein the computer code selects at least two substrings from said initial character strings by an alignment-based selection.

155. (new) The computer program product of claim 78, wherein the computer code selects at least two substrings from said initial character strings by a frequency-biased selection.

156. (new) The method of claim 90, wherein adding the product strings to a data structure comprises adding more than one product strings to the data structure.

157. (new) The method of claim 90, wherein selecting at least two substrings from said initial character strings comprises random substring selection.

158. (new) The method of claim 90, wherein selecting at least two substrings from said initial character strings comprises uniform substring selection.

159. (new) The method of claim 90, wherein selecting at least two substrings from said initial character strings comprises motif-based selection.

160. (new) The method of claim 90, wherein selecting at least two substrings from said initial character strings comprises alignment-based selection.

161. (new) The method of claim 90, wherein selecting at least two substrings from said initial character strings comprises frequency-biased selection.

162. (new) The computer program product of claim 103, wherein the computer code adds the product strings to a data structure by adding more than one product strings to the data structure.

163. (new) The computer program product of claim 103, wherein the computer code selects at least two substrings from said initial character strings by a random substring selection.

164. (new) The computer program product of claim 103, wherein the computer code selects at least two substrings from said initial character strings by a uniform substring selection.

165. (new) The computer program product of claim 103, wherein the computer code selects at least two substrings from said initial character strings by a motif-based selection.

166. (new) The computer program product of claim 103, wherein the computer code selects at least two substrings from said initial character strings by an alignment-based selection.

167. (new) The computer program product of claim 103, wherein the computer code selects at least two substrings from said initial character strings by a frequency-biased selection.

168. (new) The method of claim 116, wherein adding the product strings to a data structure comprises adding more than one product strings to the data structure.

169. (new) The method of claim 116, wherein selecting at least two substrings from said initial character strings comprises random substring selection.

170. (new) The method of claim 116, wherein selecting at least two substrings from said initial character strings comprises uniform substring selection.

171. (new) The method of claim 116, wherein selecting at least two substrings from said initial character strings comprises motif-based selection.

172. (new) The method of claim 116, wherein selecting at least two substrings from said initial character strings comprises alignment-based selection.

173. (new) The method of claim 116, wherein selecting at least two substrings from said initial character strings comprises frequency-biased selection.

174. (new) The computer program product of claim 124, wherein the computer code adds the product strings to a data structure by adding more than one product strings to the data structure.

175. (new) The computer program product of claim 124, wherein the computer code selects at least two substrings from said initial character strings by a random substring selection.

176. (new) The computer program product of claim 124, wherein the computer code selects at least two substrings from said initial character strings by a uniform substring selection.

177. (new) The computer program product of claim 124, wherein the computer code selects at least two substrings from said initial character strings by a motif-based selection.

178. (new) The computer program product of claim 124, wherein the computer code selects at least two substrings from said initial character strings by an alignment-based selection.

179. (new) The computer program product of claim 124, wherein the computer code selects at least two substrings from said initial character strings by a frequency-biased selection.